# Phenotyping Austrian Pine for Resistance Using Fourier-Transform Infrared Spectroscopy

### By Anna O. Conrad, Caterina Villari, Patrick Sherwood, and Pierluigi (Enrico) Bonello

**Abstract.** Austrian pine (*Pinus nigra*) is a valuable component of the urban landscape in the Midwestern USA. In this area, it is impacted by the fungal pathogen *Diplodia sapinea*, which causes a tip blight and canker on infected trees. While the disease can be managed through the application of fungicides and/or by preventing environmental conditions that are favorable for the pathogen, these practices only temporarily alleviate the problem. A more sustainable solution is to use resistant trees. The objective of this study was to evaluate whether Fourier-transform infrared (FT-IR) spectroscopy combined with chemometric analysis can distinguish between trees that vary in susceptibility to *D. sapinea*. Trees were phenotyped for resistance to *D. sapinea* by artificially inoculating shoots and measuring ensuing lesions seven days following inoculation. Then, three different chemometric approaches, including a type of machine learning called support vector machine (SVM), were used to evaluate whether or not trees that varied in susceptibility could be distinguished. Trees that varied in susceptibility could be discriminated based on FT-IR spectra collected prior to pathogen infection using the three chemometric approaches: soft independent modeling of class analogy, partial least squares regression, and SVM. While further validation of the predictive models is needed, the results suggest that the approach may be useful as a tool for screening and breeding Austrian pine for resistance to *D. sapinea*. Furthermore, this approach may have wide applicability in other tree/plant pathosystems of concern and economic value to the nursery and ornamental industries.

**Keywords.** *Diplodia sapinea*; Diplodia tip blight; FT-IR; *Pinus nigra*; Predictive Modeling.

## INTRODUCTION

Austrian pine (*Pinus nigra*) is an ecologically and economically important conifer tree species with a natural distribution around the Mediterranean basin. It is separated into six different subspecies spanning from western North Africa and the Iberian peninsula to the west, to Turkey and all the way into the Crimean peninsula to the east (Richardson and Rundel 1998). The type *P. nigra* subsp. *nigra* is mainly present down the Italian peninsula and spans the region that includes the Italian/Austrian Alps and the Dalmatian region (Richardson and Rundel 1998). *Pinus nigra* subsp. *nigra* and *P. nigra* subsp. *pallasiana* (mainly from Turkey) are widely planted in North America, particularly in the Midwestern USA, both as ornamental trees due to their deep green foliage and as windbreaks along major roadways due to desirable traits like deicing salt tolerance.

Like all trees, Austrian pine is susceptible to several fungal diseases, particularly under drought stress,

and in the Midwestern USA the most significant disease is Diplodia tip blight and canker, caused by the ascomycete *Diplodia sapinea* (Fr.) Fuckel 1870 (syn. *Diplodia pinea* [Desm.] Kickx., *Sphaeropsis sapinea* [Fr.: Fr.][Dyko & Sutton]). In greenhouse experiments, drought stressed Austrian pine were more susceptible to *D. sapinea* following artificial inoculation of the shoots (Sherwood et al. 2015), and the effects of water-stress on the severity of symptoms associated with *D. sapinea* infection have also been demonstrated in red pine (*P. resinosa*)(Blodgett et al. 1997a). Diplodia tip blight and canker rarely kills trees in ornamental settings, although it is known to do so in both nursery and plantation settings, for example in red pine in the North Central states (Stanosz and Carlson 1996; Haugen and Ostry 2013). However, this pathogen often causes severe disfigurement that makes the ornamental use of Austrian and other susceptible pine species, such as Scots pine (*P. sylvestris*), problematic or impossible. Oftentimes,

such disfigured trees must be removed from the landscape well before the end of their expected service life.

Integrated management for Diplodia tip blight and canker in ornamental settings includes the use of fungicides during the growing season and the alleviation of environmental conditions that favor the disease, such as insufficient water availability (Blodgett et al. 1997b; Stanosz et al. 2001). Both of these measures are only temporary and must be repeated on a regular basis, which can be costly and impractical, especially when dealing with larger size trees. A much better option would be to plant resistant trees (Boyd et al. 2013), i.e., those that have the ability to limit initial infection as well as tissue invasion by the pathogen once infected and hence express disease symptoms at much lower levels.

In other canker diseases, e.g., those produced by the pathogenic ascomycete *Fusarium circinatum* (cause of pitch canker on many pine species) and oomycete *Phytophthora ramorum* (cause of sudden oak death, a canker disease on oak and tanoak species), pathogen resistance can be assessed by measuring the length of lesions produced by artificial inoculation of the respective tree species. For example, it has been determined that if lesions produced by *F. circinatum* in Monterey pine (*P. radiata*) are limited by the host, the branches never become girdled and therefore the trees can be considered resistant (Gordon et al. 1998). Similarly, our group has shown that resistant coast live oak (*Quercus agrifolia*) can be identified by measuring the extent of lesions produced by *P. ramorum* (McPherson et al. 2014; Conrad et al. 2019).

Access to Diplodia-resistant Austrian pine would be a major advantage for the ornamental industry, where both producers and end users would benefit significantly. While inoculation-based approaches could be used to phenotype Austrian pine for resistance to *D. sapinea*, a nondestructive approach would be much more preferable, especially if it could be carried out in a considerably more rapid way.

Our group has recently demonstrated that trees can be successfully phenotyped for resistance before they become infected by using vibrational spectroscopic techniques, such as Fourier-transform infrared (FT-IR) spectroscopy, followed by chemometric analysis (i.e., multivariate statistical analysis of spectral data), in at least two pathosystems of high significance: sudden oak death of coast live oak in California and ash dieback of European ash (*Fraxinus excelsior*) in Europe (Conrad et al. 2014; Villari et al. 2018).

FT-IR spectroscopy is a chemical fingerprinting technique that measures the absorption of infrared radiation deriving from the vibration of molecular bonds in different chemical functional groups. The resulting spectra reflect the total biochemical composition of analyzed samples in a given tissue and at a given time and can be used in multivariate statistical models to identify qualitative and quantitative differences among different groups of samples. In particular, soft independent modeling of class analogy (SIMCA) can be used to classify qualitative phenotypic traits, while partial least squares regression (PLSR) can be used to classify quantitative phenotypic traits (Conrad and Bonello 2016). Support vector machine (SVM), a type of machine learning, provides another option for classifying trees based on spectral differences (Abdel-Rahman et al. 2014).

The main goal of this study was to test if this technique can be used to phenotype ornamental Austrian pine for resistance to *D. sapinea*.

## MATERIALS AND METHODS

### Plant Material

One hundred and twenty, four-year-old, open pollinated Austrian pine trees growing in one-gallon (3.7-L) plastic pots (plants kindly donated by Willoway Nursery, Madison, OH, USA) were moved from a gravel lot located on The Ohio State University (OSU) campus in Columbus, OH, to the OSU Department of Plant Pathology greenhouse on March 11, 2015. Trees were given approximately 2 L of water daily at 8:00 am and received supplemental lighting on a 14-hour regimen beginning at 6:00 am. Trees were fertilized to pot capacity with 200 ppm of Jack's Professional 20-10-20 N:P:K water-soluble fertilizer (JR Peters INC, Allentown, PA, USA) on March 12, March 23, and April 6, 2015. Trees began to break dormancy around April 6, 2015, as noted by budbreak. On April 29, 2015, 103 trees were selected for further experimentation based on plant uniformity, appearance, and bud phenological stage (i.e., selecting those currently in the elongation phase). Trees showing poor growth, trees with only a few viable shoots, and trees with shoot damage (e.g., from previous *D. sapinea* infections) were not selected.

Candles were inoculated on April 30, 2015. Using a sterile scalpel, a small wound was created at a fascicle scar (around 1 mm in size) approximately 8 cm from the shoot tip (or at the midway point on the shoot for

smaller shoots). Then, a 3-mm plug of *D. sapinea* growing on potato dextrose agar, taken from the margin of an actively growing colony, was placed mycelium-side down on the wound, and the inoculation site was sealed with parafilm M (Structure Probe, Inc., West Chester, PA, USA) to minimize contamination and desiccation. The strain of *D. sapinea* used in this experiment was the same used in previous, published work (Sherwood and Bonello 2013) and was originally isolated from symptomatic pine cones of an Austrian pine tree growing on the OSU campus in Columbus, OH. Each tree was inoculated on two separate shoots, even if only one shoot was eventually used for analyses. The dominant apical leader shoot was never selected, but shoots were otherwise chosen randomly. At the same time, two other noninoculated shoots were removed for spectroscopy and chemometric analysis of the constitutive chemical composition, excluding again the dominant apical shoot from the selection process. Harvested tissue was immediately frozen in liquid nitrogen and stored at –80 °C. Of the 103 trees that were artificially inoculated with *D. sapinea*, tissue from only 79 trees was analyzed with FT-IR spectroscopy due to an insufficient amount of available tissue from the excluded trees.

At the time of inoculation, stem diameter at 15 cm above the soil line and total plant height (not including the dominant apical leader shoot) were measured. The length of all inoculated and harvested shoots was measured, as was the length of the dominant apical leader shoot.

## Assessment of Relative Resistance

Seven days post inoculation, one of the inoculated shoots selected at random was excised, and the epidermis was removed to measure lesion lengths as a proxy of resistance. At this time, approximately 50% of the trees were beginning to show symptoms of infection (needle chlorosis and visible necrosis around the inoculation site).

To build the SIMCA chemometric model and for the SVM analysis, we selected only those trees corresponding to the quartiles of the lesion length distribution that represented the most susceptible and most resistant trees, i.e., the 25% with the longest lesions (quartile four; hereafter referred to as susceptible) and the 25% with the shortest lesions (quartile one; hereafter referred to as resistant), corresponding to 42 samples. To build the PLSR chemometric model, all

79 samples, comprising even those samples showing an intermediate phenotype (i.e., second and third quartile in the lesion length distribution; hereafter referred to as intermediate), were included in the analysis. In addition, a Welch's two sample *t*-test, with unequal variance, was performed to compare average lesion lengths between susceptible and resistant trees. Finally, a one-way ANOVA with Type III sum of squares was used to confirm that there were no significant differences in stem diameter or tree height between resistant, susceptible, and intermediate groups using the R package "Car" (Fox and Weisberg 2011). Assumptions of normality and homogeneity of variance were tested with the Shapiro-Wilk and Levene's tests, respectively. All statistical tests were performed in R version 3.5.2 (R Core Team 2018).

## Tissue Extraction for Chemical Fingerprinting

Shoot tissue corresponding to the position of the inoculations on inoculated shoots was excised from non-inoculated shoots harvested at the time of inoculation and ground in liquid nitrogen using a mortar and pestle to a fine and homogeneous powder. Aliquots of the ground tissue ($200 \pm 1$ mg) were transferred to individual 2-mL microcentrifuge tubes and stored at –80 °C until extraction, which was carried out according to Wrolstad (2005) with modifications (see Villari et al. 2018 for details). This extract was then used for chemical fingerprinting analysis.

## Chemical Fingerprinting

Seven µL of each purified extract were analyzed on an Excalibur 3500GX FT-IR benchtop spectrometer (Digilab, Randolph, MA, USA), equipped with a potassium bromide beamsplitter and a MIRacle triple-bounce zinc selenide crystal (Pike Technologies, Madison, WI, USA) attenuated total reflectance (ATR) accessory. Extracts were vacuum dried to dryness on the surface of the ATR crystal, and spectra were collected over a wavenumber range of 700 to 4000 cm$^{-1}$, which corresponds to the mid-infrared (mid-IR) region. Instrumental settings were as follows: resolution, 4 cm$^{-1}$; number of scans co-added per interferogram, 64. Spectra were visualized using Win-IR Pro 288 Software (Agilent Technologies, Santa Clara, CA, USA)(Villari et al. 2018). For SIMCA and PLSR, chemometric analysis was carried out using the modeling software Pirouette (v. 4.5, Infometrix Inc., Bothell, WA, USA),

and spectral data were normalized, smoothed, and transformed into their second derivative prior to analysis to remove multiplicative scatter and particle size interference, resolve overlapping peaks, and increase the signal-to-noise ratio (Savitzky and Golay 1964; Barnes et al. 1989; Conrad and Bonello 2016). Support vector machine (SVM) analysis was carried out using the R package "e1071" (Meyer et al. 2019).

The SIMCA model, built only with the 21 most susceptible and 21 most resistant trees, was optimized by initially including all data collected within the mid-IR region and by progressively excluding those regions with lower discriminating power between the most susceptible and most resistant trees, so that only those regions with high discriminating power were eventually included in the model. Incrementally refined models were each visualized in the SIMCA 3D class projection and Coomans plots (Coomans and Broeckaert 1986) and evaluated by observing clustering patterns of the different phenotypes. At the same time, outliers identified visually and using the outliers diagnostics plot were excluded. This included spectra with lower absorbance, which may have resulted from technical variation associated with the collection of individual FT-IR spectra and/or the extract itself. The total number of samples included in the final model was 29 (18 resistant and 11 susceptible).

The PLSR model, built on all 79 samples, was optimized with a similar strategy: visualization of the spectra and evaluation of the incrementally refined models was performed on the loading and scores plots, and outliers were identified as in Wilkerson et al. (2013). Model performance was also evaluated based on the outlier diagnostics, number of factors included, and leave-one-out cross-validation (Wilkerson et al. 2013; Conrad et al. 2014). The final model included 56 spectra (16 resistant, 30 intermediate, and 10 susceptible).

The SVM model, built using scaled raw spectral data collected across the mid-IR spectrum from the most resistant and most susceptible trees ($N = 42$), was optimized by using 10-fold cross-validation to identify the cost parameter that minimized the error rate. The cost parameter defines the size of the margin separating resistant from susceptible trees in the SVM model; models with smaller cost parameters have decreased tolerance for misclassified samples in the training data set (James et al. 2014). Due to the large number of spectral variables relative to the number of trees (biological replicates), a linear kernel was used (James et al. 2014). To assess model performance, 10-fold and 5-fold cross-validation were performed, and the accuracy and area under the receiver operating characteristic (ROC) curve from prediction scores were calculated using the R packages "e1071," "MLmetrics," and "ROCR," respectively (Sing et al. 2005; Yan 2016; Meyer et al. 2019).

## RESULTS

### Relative Resistance

Lesion lengths varied quantitatively, ranging from basically no lesion to lesions that were almost 80 mm in length (Figure 1). Trees belonging to the bottom and the top quartiles of lesion lengths (first and fourth quartiles, i.e., resistant and susceptible, respectively) were selected to build qualitative chemometric models. The average lesion lengths of resistant and susceptible trees were highly, significantly different, with susceptible trees having mean lesions approximately six times as long as resistant trees (Figure 2)(Welch's two sample $t$-test, $t = -16.0$, $P < 0.0001$, $DF = 21.8$). No significant differences in stem diameter or tree height were detected between resistant, susceptible, or intermediate trees (one-way ANOVA, $P > 0.05$).

### Chemometrics

#### Soft Independent Modeling of Class Analogy (SIMCA)

SIMCA analysis of transformed FT-IR spectra—normalized, smoothed (35 points), and with the second derivative (35 points) transformed—in the range of 698.6 to 1868.0 cm$^{-1}$ could be used to distinguish between resistant trees, those with the smallest lesion lengths (first quartile), and susceptible trees, those with the largest lesion lengths (fourth quartile)(Figures 3 and 4). FT-IR spectra were collected from extracts of noninoculated shoots and thus reflect constitutive or preinfection composition and levels of plant chemicals. The model correctly classified 89.7% of trees in the trimmed data set ($N = 29$), with 100% ($N = 18$) of resistant trees correctly classified, and 72.7% ($N = 11$) of susceptible trees correctly classified, with an interclass distance of 2.07. The larger the interclass distance, the less likely the model is to classify a sample as both resistant and susceptible.
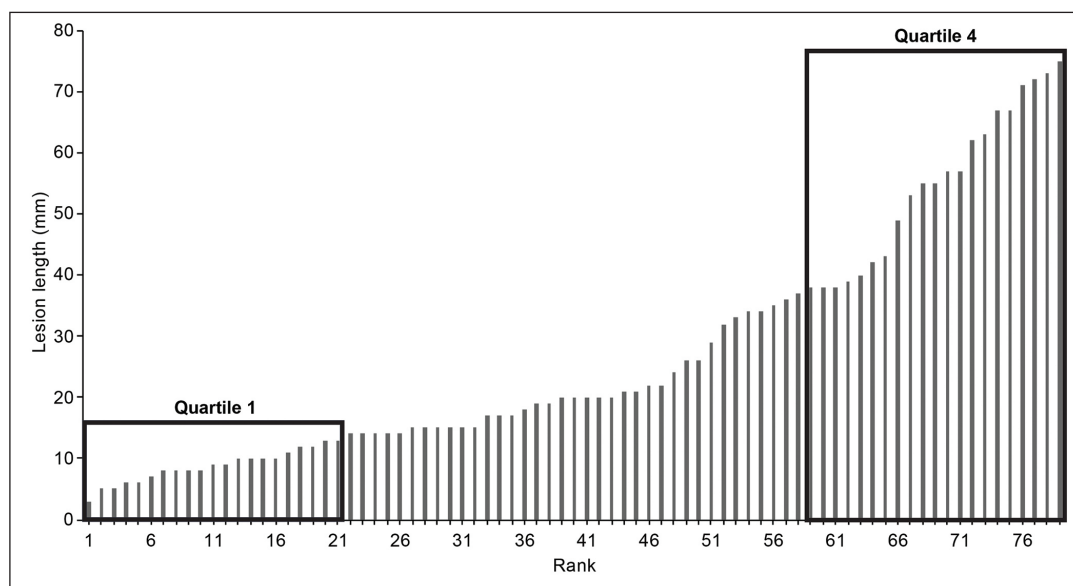
**Figure 1. Distribution of lesion lengths on Austrian pine shoots measured seven days following inoculation with *Diplodia sapinea*. Trees belonging to the first and fourth quartiles are marked, which correspond to resistant and susceptible groups, respectively.**
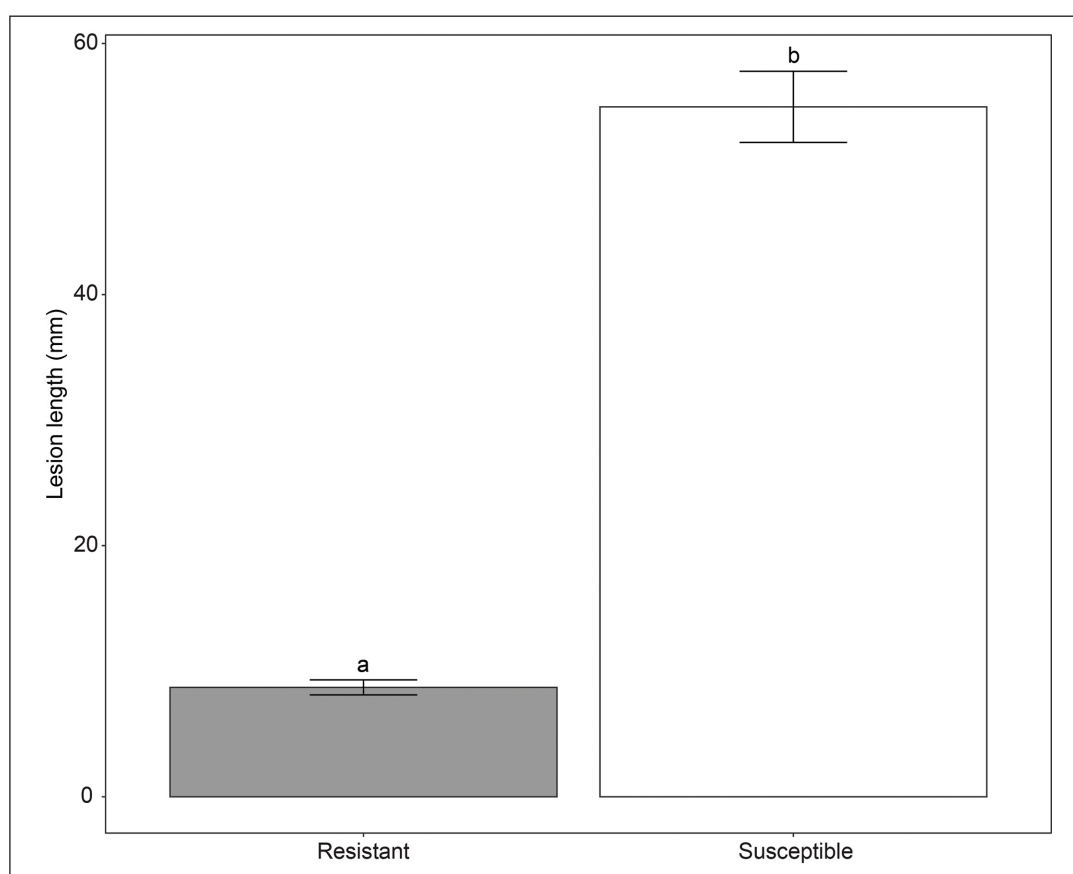


**Figure 2. Comparison of lesion lengths (± standard error) on Austrian pine shoots measured seven days following inoculation with *Diplodia sapinea*. Inoculated trees belonged to the resistant (first quartile) and susceptible (fourth quartile) groups based on lesion length distribution. Lettering denotes that resistant trees ($N = 21$) have significantly smaller lesions than susceptible trees ($N = 21$) as determined using Welch's two sample *t*-test at the $\alpha = 0.05$ level.**

## Partial Least Squares Regression (PLSR)

PLSR analysis with four factors and leave-one-out cross-validation of transformed FT-IR spectra—normalized, smoothed (25 points), and second derivative (25 points) transformed—from 802.8 to 1852.6 cm$^{-1}$ could be used to predict the length of *D. sapinea* lesions on Austrian pine ($N = 56$)(Figure 5). The standard error of cross-validation (SECV) was 12.23 mm, and the correlation coefficient of cross-validation ($r_{val}$) was 0.60, supporting a positive relationship between measured and predicted (based on FT-IR spectra) lesion lengths.

## Support Vector Machine

Ninety-three percent of analyzed trees ($N = 42$) were correctly classified as susceptible and resistant using SVM (Table 1, Figure 6). Total accuracy with 10-fold and 5-fold cross-validation was also calculated and was 64.3% and 66.7%, respectively.

## DISCUSSION

Fourier-transform infrared spectroscopy combined with chemometrics shows great promise as an alternative tool for phenotyping trees for disease resistance and could provide a more rapid and high-throughput method for use in tree breeding programs, particularly in the ornamental and nursery industries. In this study, we show that the tool can be used to distinguish between Austrian pine trees that varied in susceptibility to *D. sapinea*, the causal agent of Diplodia tip blight, using three different chemometric approaches: SIMCA, PLSR, and SVM. The latter approach uses machine learning, a form of artificial intelligence, to identify the most optimal model parameters for distinguishing between groups, specifically resistant and susceptible trees (Cortes and Vapnik 1995; Singh et al. 2016).

Inoculated Austrian pine trees showed varying levels of susceptibility (i.e., varying lesion lengths) to *D. sapinea*, indicative of a quantitative resistance response. Since susceptibility varied quantitatively, trees were separated into resistant (quartile one, smallest lesion lengths), intermediate (quartiles two and three, intermediate lesion lengths), and susceptible (quartile four, largest lesion lengths) groups. To develop classification-based predictive models for disease resistance, we focused on the tail ends of the lesion length distribution—resistant and susceptible trees. We have used this approach previously, with success, in other forest pathosystems (Conrad et al. 2014; Villari et al. 2018). Using SIMCA and SVM we were able to distinguish between these two groups.
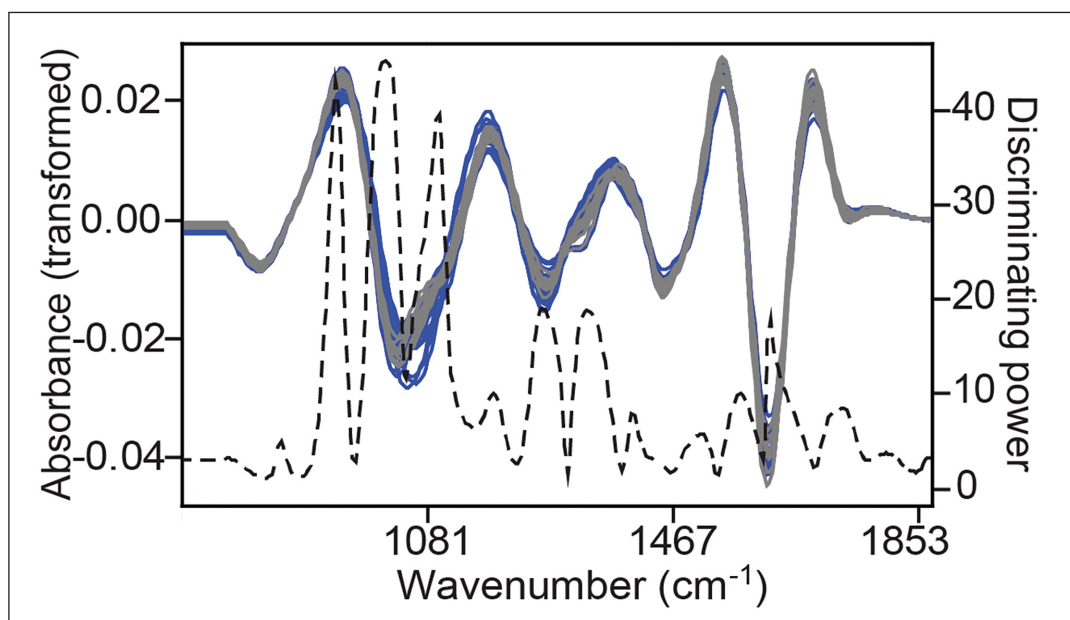


**Figure 3.** Transformed spectra from 698.6 to 1868.0 cm$^{-1}$ (left axis: blue = resistant trees and gray = susceptible trees) overlaid with SIMCA discriminating power (right axis: dashed line)($N$ = 29). Spectral features with higher discriminating powers are more important for distinguishing between resistant and susceptible trees.
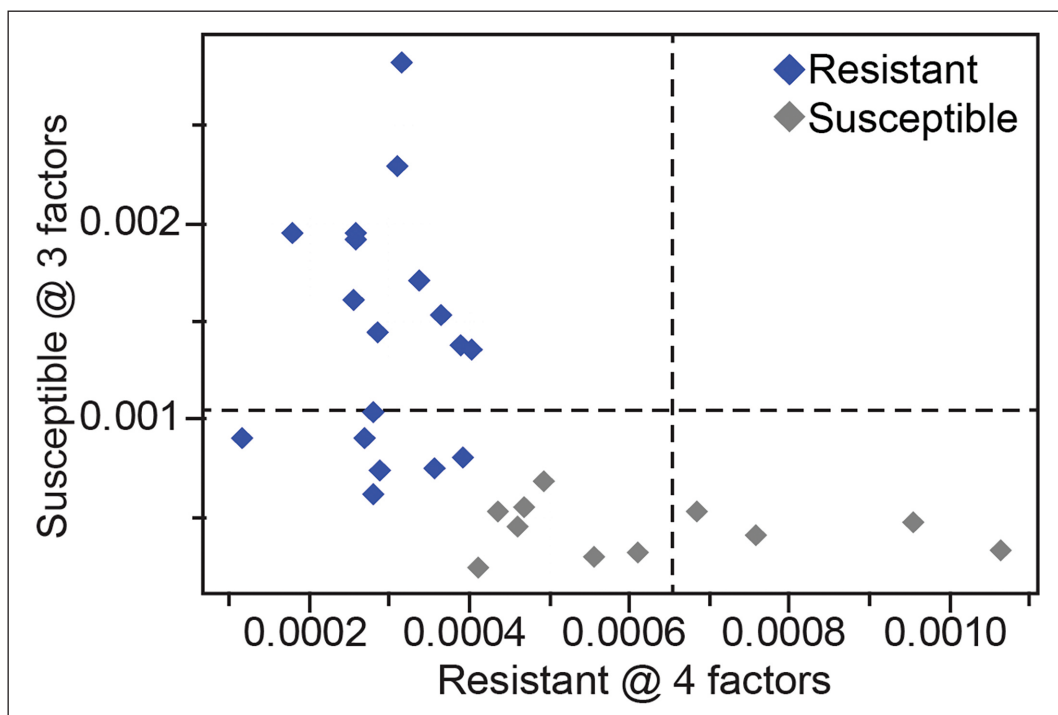
**Figure 4.** Coomans (i.e., class distances) plot showing the dimensionless distance between each sample and class—resistant and susceptible. Dashed lines indicate critical sample residual thresholds. Blue diamonds = resistant trees ($N$ = 18). Gray diamonds = susceptible trees ($N$ = 11). Samples in the top left quadrant are always classified as resistant. Samples in the bottom right quadrant are always classified as susceptible. Samples in the bottom left quadrant could be either resistant or susceptible. Every sample could be classified by the model, so the top right quadrant is empty.
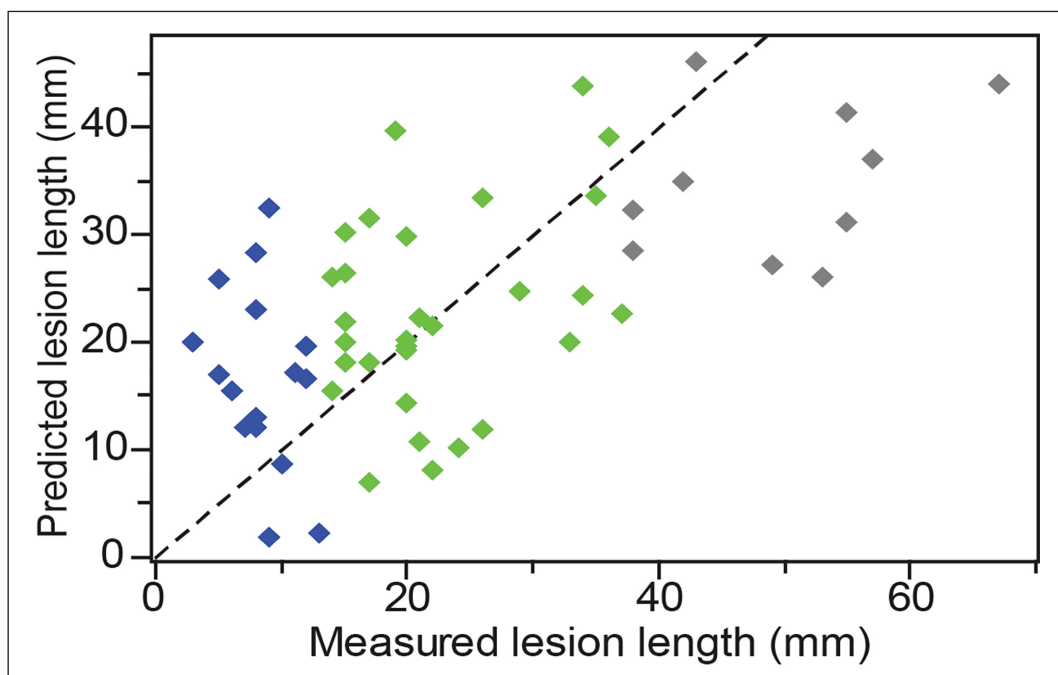


**Figure 5.** Measured versus predicted lesion lengths (in mm) based on PLSR analysis of spectral data from 802.8 to 1852.6 cm[-1]. Blue diamonds = resistant trees ($N$ = 16). Green diamonds = intermediate trees ($N$ = 30). Gray diamonds = susceptible trees ($N$ = 10). The correlation coefficient of cross-validation ($r_{val}$) is 0.60.

However, a non-negligible number of trees had to be trimmed from the final, optimized SIMCA model (13 out of a total $N = 42$), most of which belonged to the susceptible group. This suggests that there was more variation in the FT-IR spectra of trees classified as susceptible compared to resistant, which may have impacted the classification performance, particularly that of susceptible trees, and level of bias (due to outlier removal) of the SIMCA model. In contrast, with SVM, 100% of trees ($N = 42$) were included in the final, optimized model. Model accuracy was comparable, in particular for resistant trees, in both cases: 89.7% ($N = 29$) and 92.9% ($N = 42$) of trees in total were correctly classified in the SIMCA and SVM models, respectively. Taken together, these results suggest that SVM is a more desirable approach for the development of classification-based disease predictive models using FT-IR spectra, although the analysis of additional biological samples may help to improve the classification performance of both models. Ultimately, the level of allowable misclassification or the classification threshold should be based on the goals of the disease management and/or resistance screening program.

In this study, PLSR was also used to predict susceptibility of trees, via lesion length, based on pre-infection FT-IR spectra. This method is an alternative to classification-based predictive approaches and has been used extensively in the literature, for example to predict the susceptibility of eucalyptus (*Eucalyptus*

**Table 1. Results of support vector machine (SVM) analysis to classify trees as resistant or susceptible based on scaled spectral data collected from 700–4000 cm[-1].**

| Group | Correctly classified (%) | Misclassified (%) | *N* | Accuracy[2] | AUC[3] |
|---|---|---|---|---|---|
| Resistant | 95.2 | 4.8 | 21 | | |
| Susceptible | 90.5 | 9.5 | 21 | | |
| Model[1] | | | 42 | 0.929 | 0.929 |

[1]Linear kernel support vector machine with cost = 1. [2]Accuracy classification score. [3]AUC: area under the receiver operating characteristic curve from prediction scores.
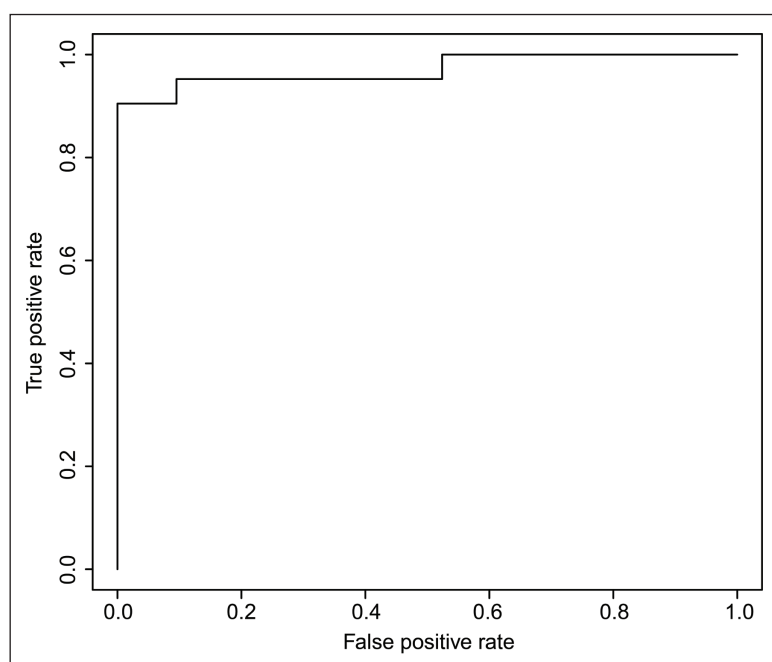


**Figure 6. Receiver operating characteristic (ROC) curve for support vector machine analysis showing the relationship between the false positive rate and the true positive rate. The area under the curve (AUC) is 0.929.**

*grandis*) to *Leptocybe invasa*, a gall wasp (Naidoo et al. 2018), and to predict decay resistance in Scots pine (*P. sylvestris*) to *Poria placenta*, a brown rot fungus (Flæte and Haartveit 2004). As with SIMCA, some trees had to be trimmed from the final, optimized PLSR model (29% of $N = 79$), although in contrast to SIMCA and SVM, intermediate trees were also included. This suggests that the technique is also capable of detecting minor quantitative differences in susceptibility and not only differences between extreme groups (in our case, resistant versus susceptible).

Since the number of trees included in this study was limited, further evaluation and validation of predictive models is needed before the approach can be implemented widely and on a production scale. Nonetheless the combination of FT-IR spectroscopy with chemometrics shows great promise as a tool for resistance screening efforts and may be widely applicable in the ornamental and nursery industries in the future. In support of this statement, FT-IR spectroscopy has been used to differentiate between elms that differed in susceptibility to Dutch elm disease (Martin et al. 2005) and in our own work could be used to distinguish between resistant and susceptible coast live oak to sudden oak death (Conrad et al. 2014) and European ash to ash dieback (Villari et al. 2018).

This approach provides an alternative to traditional, inoculation-based methods for screening Austrian pine for resistance to *D. sapinea* and has the potential to allow for more rapid and high-throughput phenotyping in the future. Furthermore, the approach may be useful not only for trees, but for other woody perennials, such as boxwood (*Buxus* spp.), which is greatly impacted by boxwood blight, a disease caused by the fungus *Calonectria pseudonaviculata*. In this case, an FT-IR and chemometrics approach could be used to supplement efforts aimed at assessing the susceptibility of boxwood cultivars (e.g., Guo et al. 2016) and for future resistance breeding programs. Therefore, the approach may have wide applicability in the ornamental and nursery industries, since identifying disease resistant trees and plants for use in the urban landscape is critical to the economic success of those industries.

## LITERATURE CITED

Abdel-Rahman EM, Mutanga O, Adam E, Ismail R. 2014. Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS Journal of Photogrammetry Remote Sensing*. 88:48-59.

Barnes RJ, Dhanoa MS, Lister SJ. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*. 43:772-777.

Blodgett JT, Kruger EL, Stanosz GR. 1997a. Effects of moderate water stress on disease development by *Sphaeropsis sapinea* on red pine. *Phytopathology*. 87:422-428.

Blodgett JT, Kruger EL, Stanosz GR. 1997b. *Sphaeropsis sapinea* and water stress in a red pine plantation in central Wisconsin. *Phytopathology*. 87:429-434.

Boyd IL, Freer-Smith PH, Gilligan CA, Godfray HCJ. 2013. The consequence of tree pests and diseases for ecosystem services. *Science*. 342:1235773.

Conrad AO, Bonello P. 2016. Application of infrared and Raman spectroscopy for the identification of disease resistant trees. *Frontiers in Plant Science*. 6:1-8.

Conrad AO, McPherson BA, Lopez-Nicora HD, D'Amico KM, Wood DL, Bonello P. 2019. Disease incidence and spatial distribution of host resistance in a coast live oak/sudden oak death pathosystem. *Forest Ecology and Management*. 433:618-624.

Conrad AO, Rodriguez-Saona LE, McPherson BA, Wood DL, Bonello P. 2014. Identification of *Quercus agrifolia* (coast live oak) resistant to the invasive pathogen *Phytophthora ramorum* in native stands using Fourier-transform infrared (FT-IR) spectroscopy. *Frontiers in Plant Science*. 5:521.

Coomans D, Broeckaert I. 1986. *Potential pattern recognition in chemical and medical decision making*. Letchworth (England): Research Studies Press LTD. 256 p.

Cortes C, Vapnik V. 1995. Support-vector networks. *Machine Learning*. 20:273-297.

Flæte PO, Haartveit EY. 2004. Non-destructive prediction of decay resistance of *Pinus sylvestris* heartwood by near infrared spectroscopy. *Scandinavian Journal of Forest Research*. 19:55-63.

Fox J, Weisberg S. 2011. *car: An {R} companion to applied regression*. 2nd Ed. Thousand Oaks (CA, USA): Sage. 472 p.

Gordon TR, Wikler KR, Clark SL, Okamoto D, Storer AJ, Bonello P. 1998. Resistance to pitch canker disease, caused by *Fusarium subglutinans* f.sp. *pini* in Monterey pine. *Plant Pathology*. 47:706-711.

Guo Y, Olsen RT, Kramer M, Pooler M. 2016. Use of mycelium and detached leaves in bioassays for assessing resistance to boxwood blight. *Plant Disease*. 100:1622-1626.

Haugen LM, Ostry ME. 2013. Long-term impact of shoot blight disease on red pine saplings. *Northern Journal of Applied Forestry*. 30:170-174.

James G, Witten D, Hastie T, Tibshirani R. 2014. *An introduction to statistical learning: with applications in R*. New York (NY, USA): Springer. 426 p.

Martin J, Solla A, Woodward S, Gil L. 2005. Fourier transform-infrared spectroscopy as a new method for evaluating host resistance in the Dutch elm disease complex. *Tree Physiology*. 25:1331-1338.

McPherson BA, Mori SR, Opiyo SO, Conrad AO, Wood DL, Bonello P. 2014. Association between resistance to an introduced invasive pathogen and phenolic compounds that may serve as biomarkers in native oaks. *Forest Ecology and Management*. 312:154-160.

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. 2019. e1071: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package. https://cran.r-project.org/package=e1071

Naidoo S, Christie N, Acosta JJ, Mphahlele MM, Payn KG, Myburg AA, Külheim C. 2018. Terpenes associated with resistance against the gall wasp, *Leptocybe invasa*, in *Eucalyptus grandis*. *Plant, Cell & Environment*. 41:1840-1851.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Richardson DM, Rundel PW. 1998. Ecology and biogeography of *Pinus*: an introduction. p. 3-46. In: Richardson DM, editor. *Ecology and biogeography of* Pinus. Cambridge (UK): Cambridge University Press. 527 p.

Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*. 36:1627-1639.

Sherwood P, Bonello P. 2013. Austrian pine phenolics are likely contributors to systemic induced resistance against *Diplodia pinea*. *Tree Physiology*. 33:845-854.

Sherwood P, Villari C, Capretti P, Bonello P. 2015. Mechanisms of induced susceptibility to *Diplodia* top blight in drought-stressed Austrian pine. *Tree Physiology*. 35:549-562.

Sing T, Sander O, Beerenwinkel N, Lengauer T. 2005. ROCR: visualizing classifer performance in R. *Bioinformatics*. 21:7881.

Singh A, Ganapathysubramanian B, Singh AK, Sarkar S. 2016. Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science*. 21:110-124.

Stanosz GR, Blodgett JT, Smith DR, Kruger EL. 2001. Water stress and *Sphaeropsis pinea* as a latent pathogen of red pine seedlings. *New Phytologist*. 149:531-538.

Stanosz GR, Carlson JC. 1996. Association of mortality of recently planted seedlings and established saplings in red pine plantations with Sphaeropsis collar rot. *Plant Disease*. 80:750-753.

Villari C, Dowkiw A, Enderle R, Ghasemkhani M, Kirisits T, Kjær ED, Marčiulynienė D, McKinney LV, Metzler B, Muñoz F, Nielsen LR, Pliūra A, Stener LG, Suchockas V, Rodriguez-Saona L, Bonello P, Cleary M. 2018. Advanced spectroscopy-based phenotyping offers a potential solution to the ash dieback epidemic. *Scientific Reports*. 8:17448.

Wilkerson ED, Anthon GE, Barrett DM, Sayajon GFG, Santos AM, Rodriguez-Saona LE. 2013. Rapid assessment of quality parameters in processing tomatoes using hand-held and bench-top infrared spectrometers and multivariate analysis. *Journal of Agriculture and Food Chemistry*. 61:2088-2095.

Wrolstad RE. 2005. *Handbook of food analytical chemistry*. Hoboken (NJ, USA): J. Wiley. 606 p.

Yan Y. 2016. MLmetrics: machine learning evaluation. R package version 1.1.1. https://CRAN.R-project.org/package=MLmetrics

## ACKNOWLEDGMENTS

*Anna O. Conrad (corresponding author)*
*Department of Plant Pathology*
*The Ohio State University*
*Columbus, OH, USA*
*conrad.245@osu.edu*
*614-688-5409*

*Caterina Villari*
*D.B. Warnell School of Forestry & Natural Resources*
*University of Georgia*
*Athens, GA, USA*

*Patrick Sherwood*
*Department of Plant Pathology*
*The Ohio State University*
*Columbus, OH, USA*
*Southern Swedish Forest Research Centre\**
*Swedish University of Agricultural Sciences*
*Alnarp, Sweden*

*Pierluigi (Enrico) Bonello*
*Department of Plant Pathology*
*The Ohio State University*
*Columbus, OH, USA*

*\*current address*

**Conflicts of Interest:** The authors reported no conflicts of interest.

**Résumé.** Le pin noir d'Autriche (*Pinus nigra*) est une composante précieuse du paysage urbain du Midwest américain. Dans cette région, il est affecté par le pathogène fongique *Diplodia sapinea*, lequel cause la brûlure des pousses et un chancre sur les arbres infectés. Bien que la maladie puisse être régie par l'application de fongicide et/ou en évitant les conditions environnementales favorables au pathogène, ces pratiques n'atténuent que temporairement le problème. Une solution plus durable consiste à utiliser des arbres résistants. L'objectif de cette recherche était d'évaluer si la spectroscopie infrarouge transformée de Fourier (FT-IR) jumelée avec l'analyse chimiométrique pouvait distinguer la variabilité de la susceptibilité des arbres à *D. sapinea*. Les arbres ont été phénotypés pour leur résistance à *D. sapinea* en inoculant artificiellement des pousses et en mesurant les lésions résultantes sept jours après l'inoculation. Trois approches chimiométriques différentes, incluant un type d'apprentissage machine appelé machine à vecteur de support (MVS), furent utilisées afin d'évaluer si les arbres dont la susceptibilité était variable, pouvaient être caractérisés ou non. La susceptibilité des arbres variait sur la base des spectres FT-IR recueillis préalablement à l'infection du pathogène en recourant aux trois approches chimiométriques: modélisation douce indépendante de l'analogie des classes, régression partielle par les moindres carrés et MVS. Tandis qu'une validation plus poussée des modèles prédictifs est nécessaire, les résultats suggèrent que l'approche peut être utile en tant qu'outil

de dépistage et de sélection du pin noir d'Autriche pour la résistance à *D. sapinea*. De plus, cette approche peut avoir une large applicabilité dans d'autres pathosystèmes arbre/plante d'intérêt et de valeur économique pour les pépinières et le milieu ornemental.

**Zusammenfassung.** Die österreichische Kiefer (*Pinus nigra*) ist ein wertvoller Bestandteil der Stadtlandschaft im Mittleren Westen der USA. In diesem Gebiet wird sie von dem Pilzerreger *Diplodia sapinea* befallen, der an infizierten Bäumen eine Kraut- und Knollenfäule verursacht. Die Krankheit kann zwar durch die Anwendung von Fungiziden und/oder durch die Verhinderung von für den Erreger günstigen Umweltbedingungen bekämpft werden, aber lindern das Problem nur vorübergehend. Eine nachhaltigere Lösung ist die Verwendung resistenter Bäume. Ziel dieser Studie war es, zu beurteilen, ob die Fourier-Transform-Infrarotspektroskopie (FT-IR) in Kombination mit einer Chemometrie-Analyse zwischen Bäumen unterscheiden kann, deren Anfälligkeit für *D. sapinea* unterschiedlich stark ausgeprägt ist. Die Bäume wurden für die Resistenz gegen *D. sapinea* phänotypisiert, indem die Triebe künstlich geimpft und die daraus resultierenden Läsionen sieben Tage nach der Inokulation gemessen wurden. Sieben Tage nach der Inokulation wurden die Bäume auf Resistenz gegen *D. sapinea* phänotypisiert, indem die Triebe künstlich inokuliert und die daraus resultierenden Läsionen gemessen wurden. Anschließend wurden drei verschiedene chemometrische Ansätze, einschließlich einer Art von maschinellem Lernen, der so genannten Support-Vektor-Maschine (SVM), verwendet, um zu beurteilen, ob Bäume mit unterschiedlicher Anfälligkeit unterschieden werden konnten oder nicht. Bäume mit unterschiedlicher Anfälligkeit konnten auf der Grundlage von FT-IR-Spektren unterschieden werden, die vor der Pathogeninfektion mit den drei chemometrischen Ansätzen gesammelt wurden: weiche unabhängige Modellierung der Klassenanalogie, partielle Regression der kleinsten Quadrate und SVM. Obwohl eine weitere Validierung der prädiktiven Modelle erforderlich ist, deuten die Ergebnisse darauf hin, dass der Ansatz als Werkzeug für das Screening und die Züchtung österreichischer Kiefern auf Resistenz gegen *D. sapinea* nützlich sein könnte. Darüber hinaus könnte dieser Ansatz eine breite Anwendbarkeit in anderen Baum-/Pflanzen-Pathosystemen haben, die für die Baumschulen und die Zierpflanzenindustrie von Belang und von wirtschaftlichem Wert sind.

**Resumen.** El pino austriaco (*Pinus nigra*) es un componente valioso del paisaje urbano en el medio oeste de EE. UU. En esta zona se ve afectado por el patógeno fúngico *Diplodia sapinea,* que causa una caída de la punta y cancros en los árboles infectados. Si bien la enfermedad se puede controlar mediante la aplicación de fungicidas y/o mediante la prevención de las condiciones ambientales que son favorables para el patógeno, estas prácticas sólo alivian temporalmente el problema. Una solución más sostenible es utilizar árboles resistentes. El objetivo de este estudio fue evaluar si la espectroscopía infrarroja de transformación de Fourier (FT-IR) combinada con el análisis quimiométrico puede distinguir entre los árboles que varían en susceptibilidad a *D. sapinea*. Los árboles fueron fenotipos para la resistencia a *D. sapinea* inoculando artificialmente brotes y midiendo las lesiones subsiguientes siete días después de la inoculación. Luego, se utilizaron tres enfoques quimiométricos diferentes, incluido un tipo de aprendizaje automático llamado máquina de vectores de soporte (SVM), para evaluar si se podían distinguir o no árboles que variaban en susceptibilidad. Los árboles que variaron en susceptibilidad podrían ser discriminados en función de los espectros FT-IR recogidos antes de la infección por patógenos utilizando los tres enfoques quimiométricos: modelado suave independiente de analogía de clase, regresión parcial de mínimos cuadrados y SVM. Si bien se necesita una mayor validación de los modelos predictivos, los resultados sugieren que el enfoque puede ser útil como herramienta para la detección y cultivo del pino austriaco para la resistencia a *D. sapinea*. Además, este enfoque puede tener una amplia aplicabilidad en otros patosistemas arbóreos/plantas de interés económico para las industrias de viveros y ornamentales.